PCC and the transition to Linked Data - implementation phase

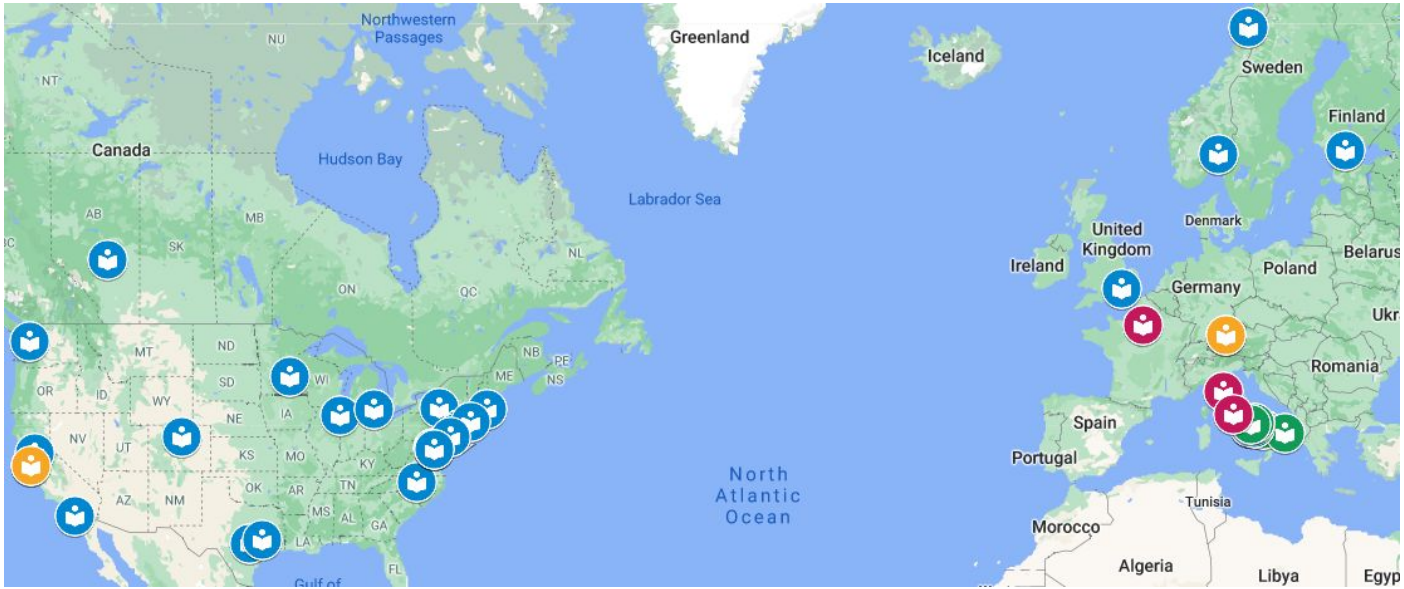
PCC Policy Committee Meeting, October 30th 2020

Michele Casalini
michele@casalini.it

https://share-vde.org
info@share-vde.org
https://wiki.share-vde.org/

Share-VDE achievements over the last year

The Share family is increasing



Most recent members:

The National Library of Finland
The British Library
The Library of Congress

The Share family participating institutions

Share-VDE Full members

Duke University
New York University
Stanford University
University of Alberta
University of Chicago
University of Michigan at Ann Arbor
University of Pennsylvania
Yale University

National Libraries

Library of Congress
National Library of Finland
National Library of Norway
The British Library

LD4P Cohort members

Cornell University
Frick Art Reference Library
Harry Ransom Center Texas A&M
Harvard University
National Library of Medicine
Northwestern University
Princeton University
University of California Davis
University of California San Diego
University Colorado at Boulder
University of Minnesota
University of Texas A&M
University of Washington

Share-Music project

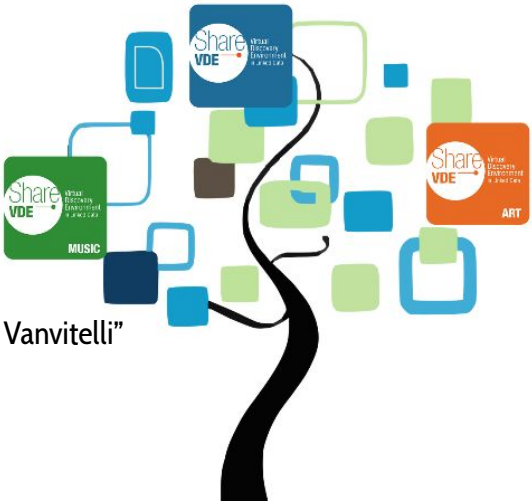
Bayerische Staatsbibliothek
Library of Congress
Stanford University

Share-Catalogue Institutions

Università degli Studi di Napoli "Federico II"
Università degli Studi della Basilicata
Università degli Studi di Napoli L'Orientale
Università degli Studi di Napoli Parthenope
Università del Salento
Università degli Studi di Salerno
Università degli Studi del Sannio RCost
Università degli Studi della Campania "Luigi Vanvitelli"

Share-Art (Kubikat-LOD) project

Max-Planck-Institut:
Kunsthistorisches Institut in Florenz
Biblioteca Hertziana Rome
Central Institute of Art History Munich
Deutsches Forum für Kunstgeschichte Paris



See also
https://wiki.share-vde.org/wiki/ShareVDE:Main_Page/SVDE_institutions

Share-VDE Advisory Council & Working Groups

Libraries members of SVDE working groups and parallel projects are constantly contributing with their Subject Matter Experts to requirements gathering, functional analysis and feedback to developments

The Share-VDE development team expanded

The SVDE IT team has been restructured and enlarged in order to cope with the increasing complexity of the developments, meet the needs of the community and interconnect with several projects. Five main development teams:

Infrastructure & Architecture

SVDE Backend (Database and Indexing, APIs development)

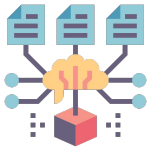
SVDE Frontend (Frontend components, APIs development, SVDE portal and J.Cricket search functions)

APIs for Penn

J.Cricket Editing functions

Main 2020 achievements #1

The design of the UI has been completed and enhanced by the activities around the J.Cricket Cluster Knowledge Base editor, the requirements for the University of Pennsylvania localisation and for the Kubikat-LOD parallel project



Backend infrastructure rearranged to respond to many layers and complex search logic → general revision of the development plans

The CKB is being enhanced with new attributes and new controlled vocabularies as a result of the UI design and the revision of the backend infrastructure

Tenant modelling: skins and tenant architecture to assure benefits from cooperation without forgetting the independence of the libraries (e.g. Penn and National Bibliography future skins; Kubikat tenant)

Main 2020 achievements #2

Design of the J.Cricket Cluster Knowledge Base Editor with the CKB Editor working group

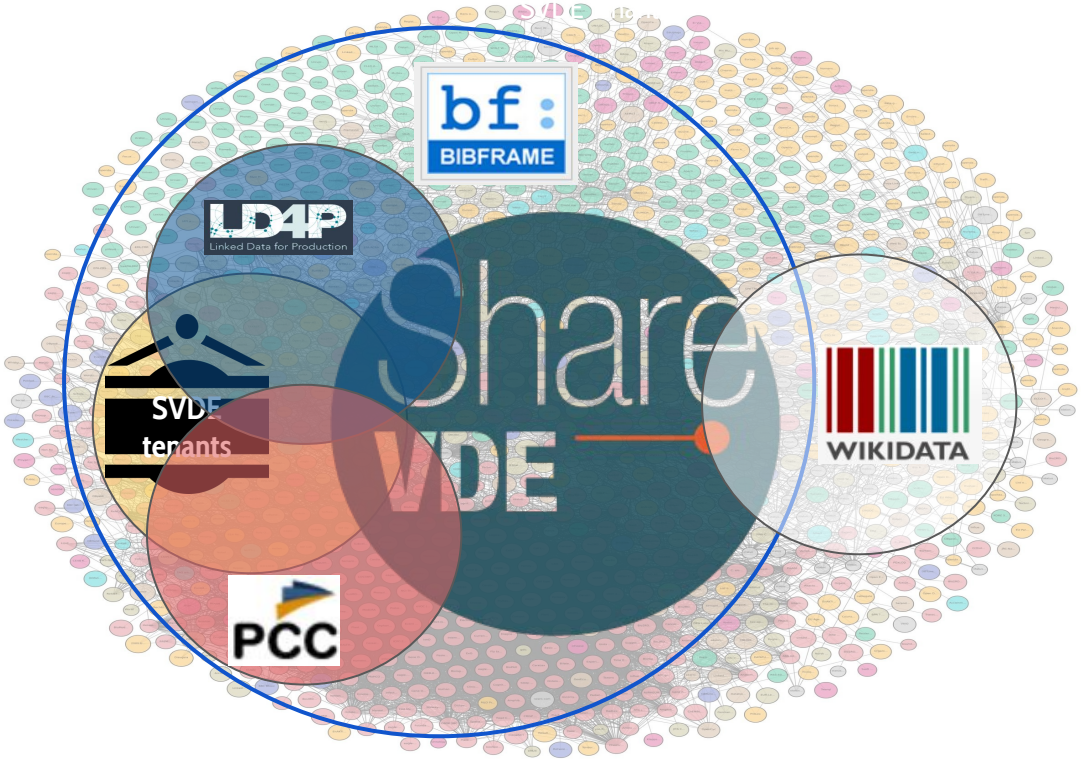
Analysis of authority services with the Authority Identifier Management Services working group and initial developments of the actual workflows

Analysis for the integration with Wikidata and ISNI in AIMS and CKB working groups

Revision of the entity model with the Sapientia Entity Identification working group:

- svde:Opus approved
- 4 layers in SVDE entity model: svde:Opus | svde:Work | svde:Instance | svde:Item
- svde:Opus and svde:Work are types of bf:Work → this ensures interoperability
- work on svde:Instance implementation

Community engagement



Communication and events:

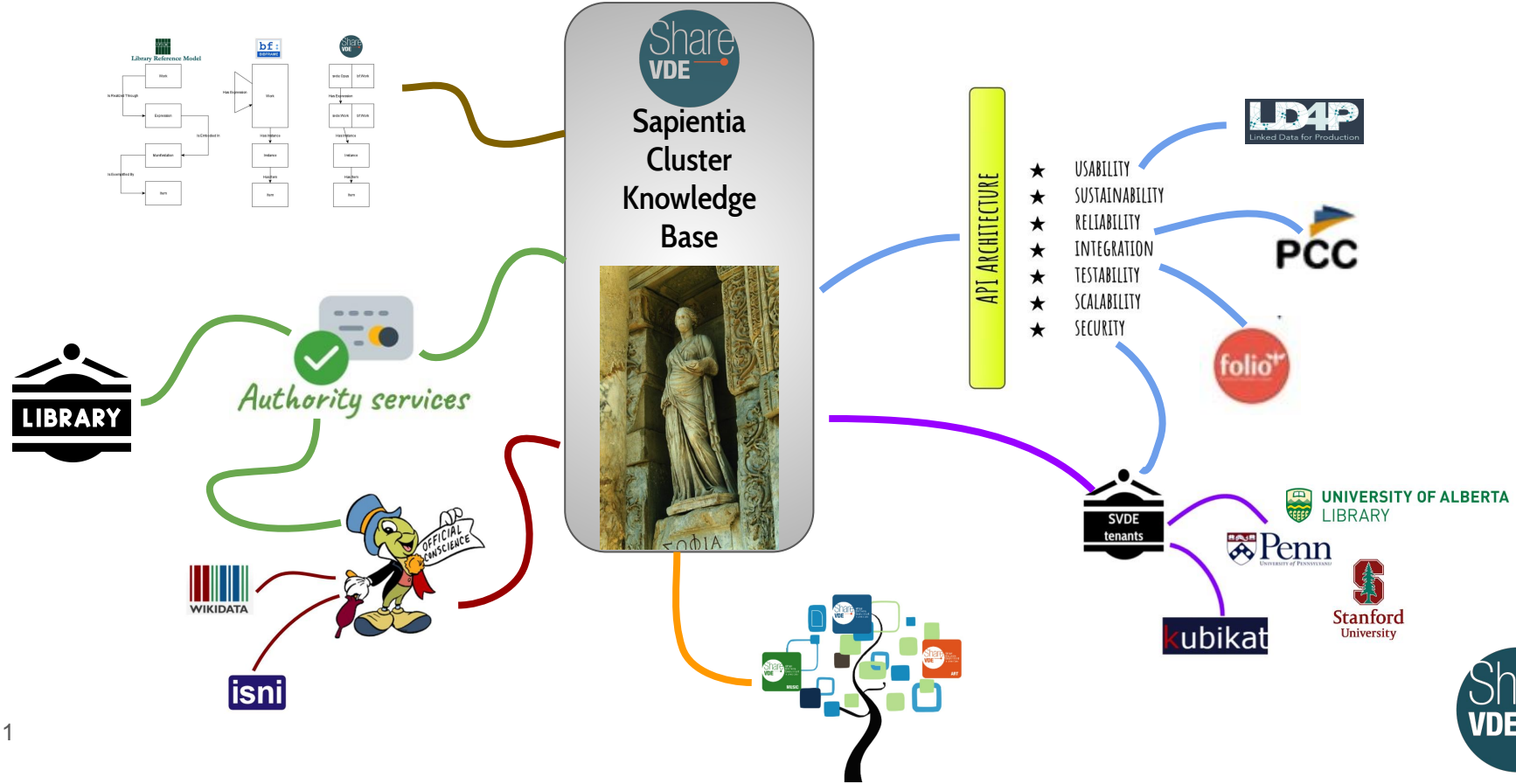
wiki platform https://wiki.share-vde.org/ that will serve both as a showcase as well as a collaborative platform among SVDE members

events: ALA MW, LIBER conference, LD4 conference, European BIBFRAME Workshop

articles and resources: https://wiki.share-vde.org/wiki/ShareVDE:Resources#Bibliographic_references

Where we are now

Share-VDE Sapientia CKB ecosystem



Developing interconnections

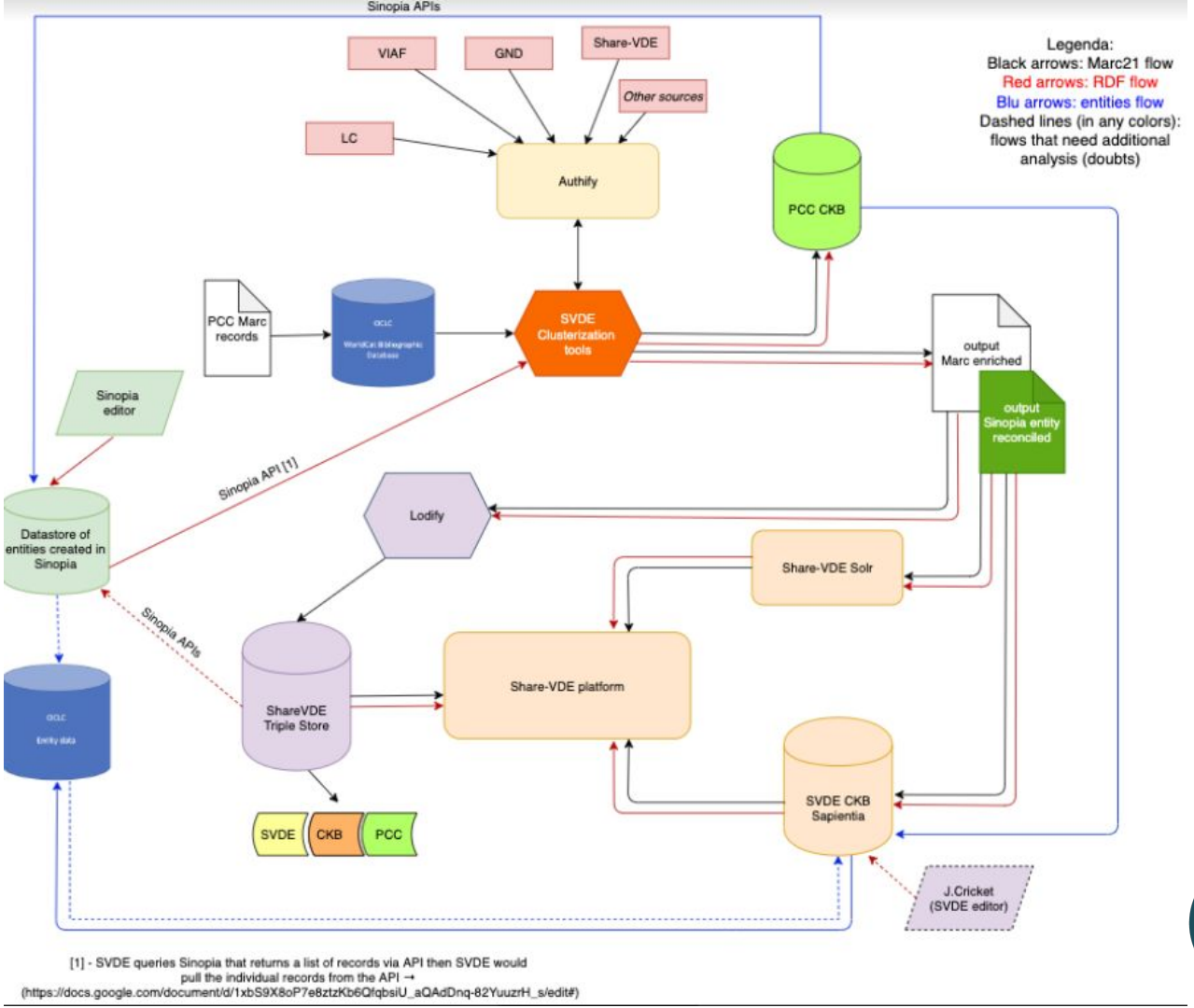
Participation in LD4P3: an API based, two-ways flow Sinopia-SVDE is being put in place:

- first step done → API pipeline that pulls records from Sinopia to SVDE
- next steps: working on interoperability of entity models and on the revision of SVDE data to be queried by Sinopia/QA

Wikidata as source of identification to be integrated in SVDE to enrich linked data entities and MARC records → connection with Casalini participation in the PCC Wikidata pilot

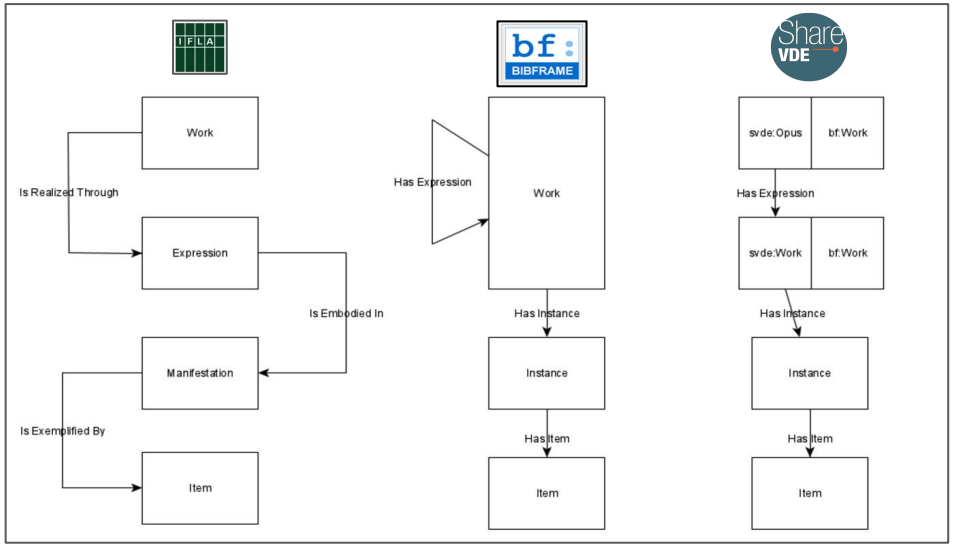
Participation in LD4P3: towards closing the loop

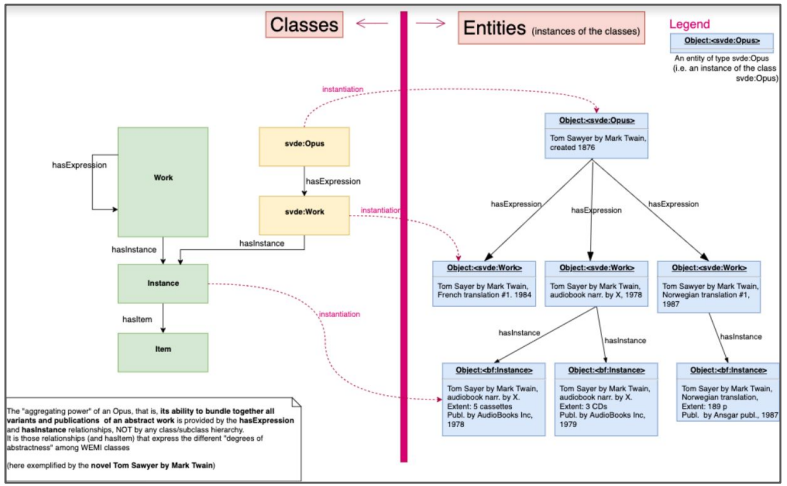
[See the diagram online here]



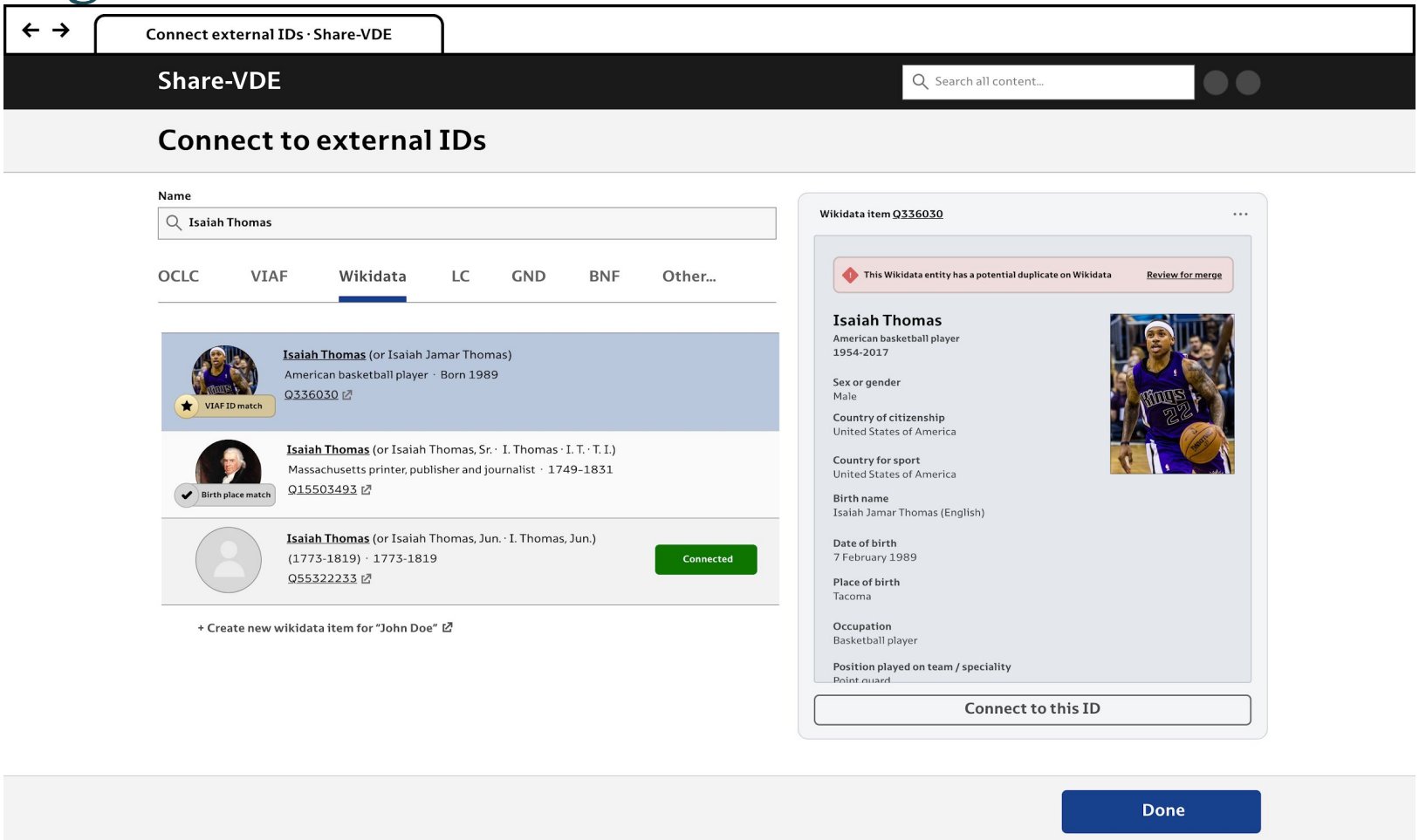
Participation in LD4P3: the challenge of data models interoperability

See the SVDE entity model compared to BF and LRM and an example of application of the model





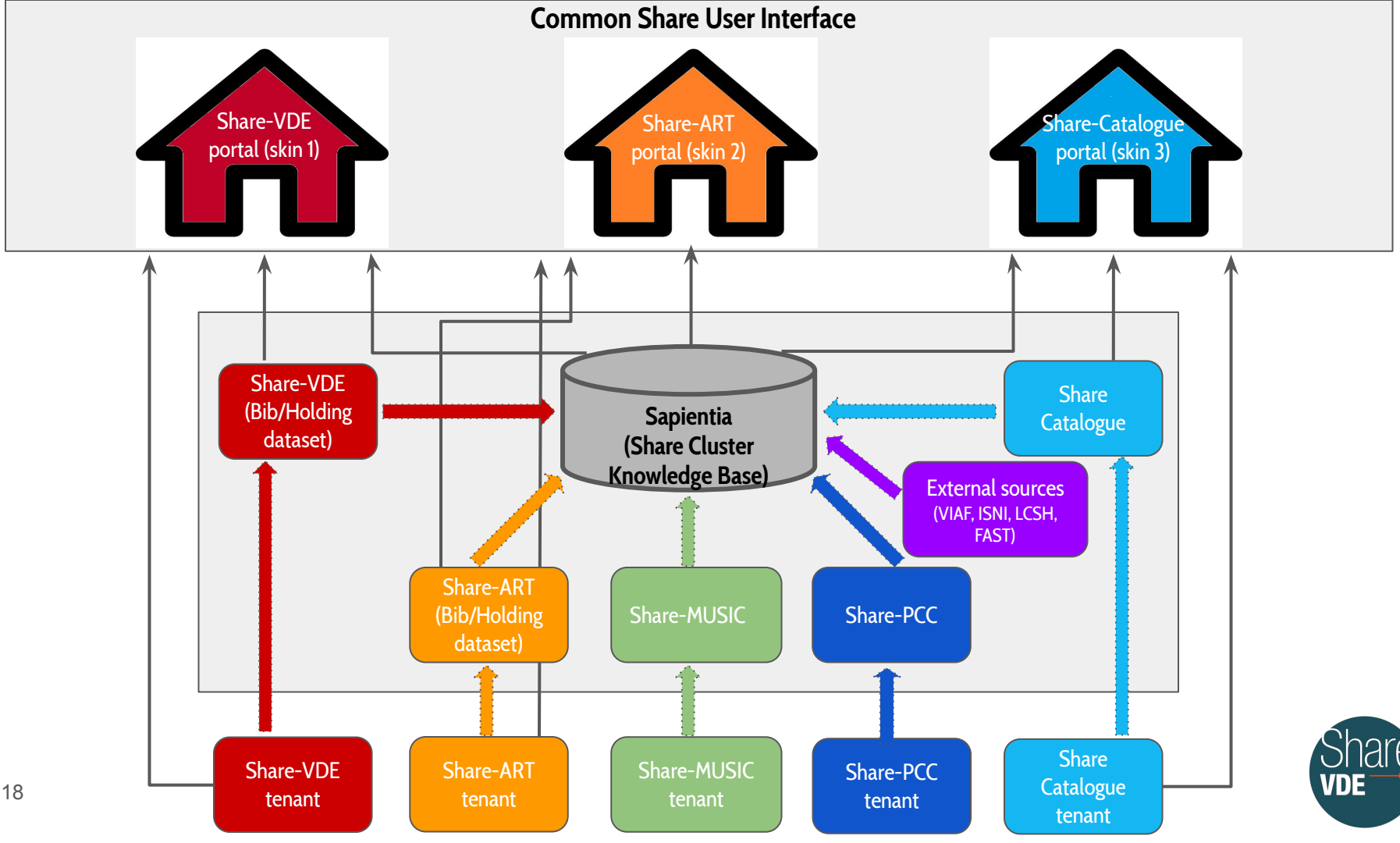
Integration of Wikidata IDs in SVDE with J.Cricket



Share-VDE and the PCC

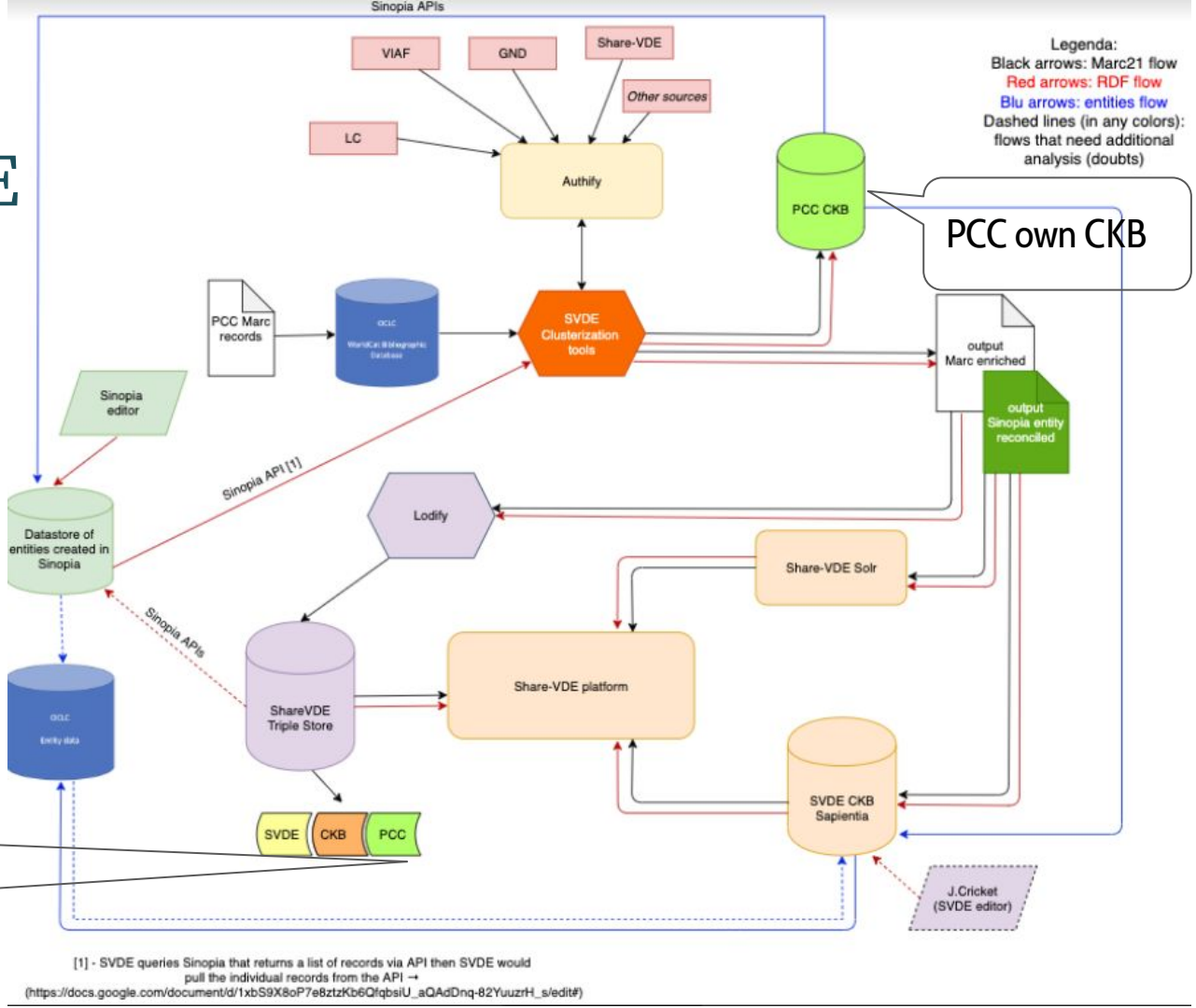
PCC as autonomous tenant

- The Sapientia Cluster Knowledge Base is shared by all tenants:
  - all tenants and external data sources feed the CKB;
- tenants can also be represented through different User Interface skins;
  - Share-VDE new UI demo video;
  - Kubikat-LOD new UI prototype;
- possible options for further future tenants:
  - tenant of national libraries (e.g. including the National Libraries of Norway and Finland, and other national libraries that will join the Share family);
  - cross-library tenants structured by domain: art tenant, music tenant etc.;
  - **—> PCC tenant with PCC libraries data.**

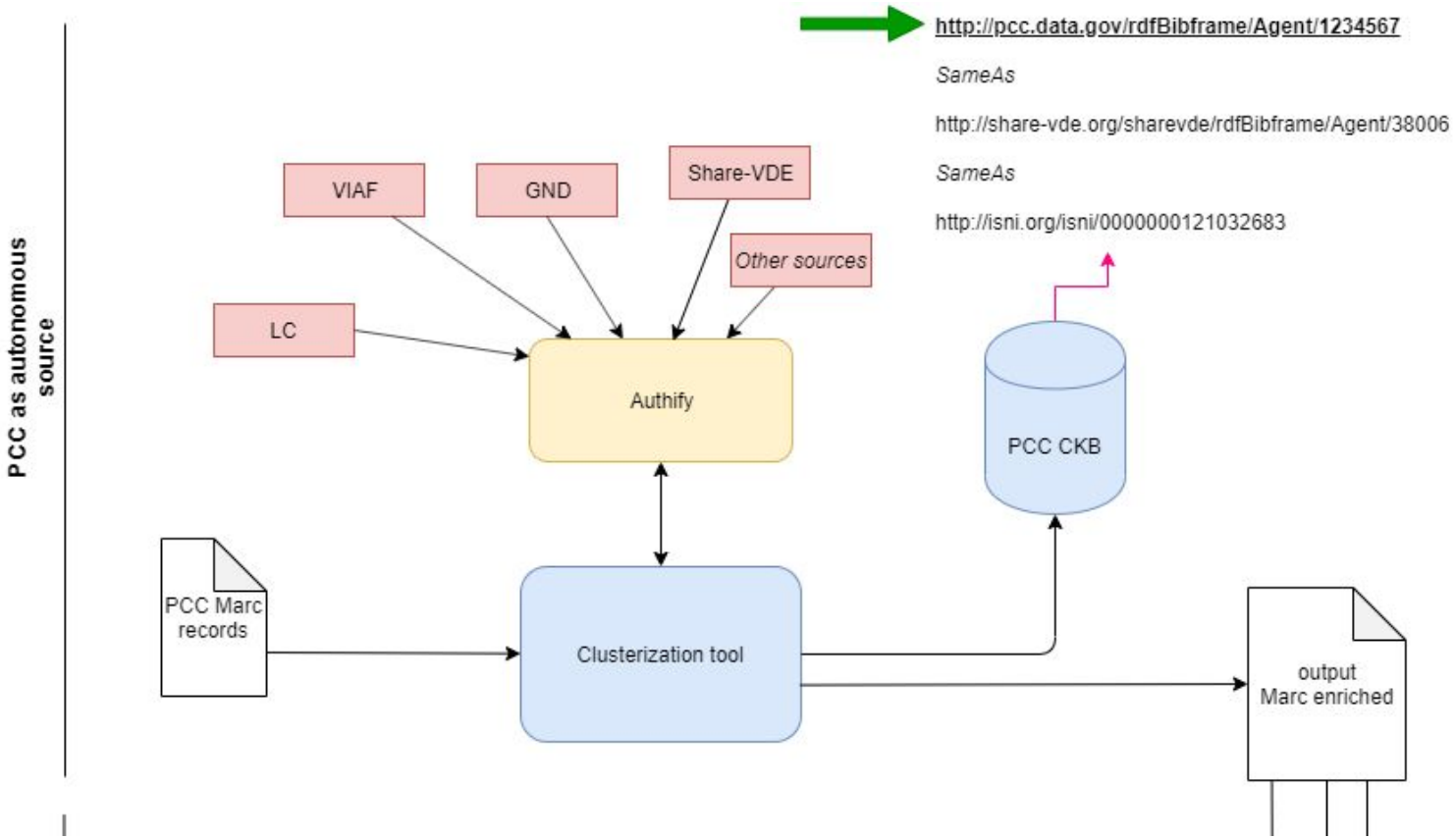

Common Share User Interface
Share-VDE portal (skin 1)
Share-ART portal (skin 2)
Share-Catalogue portal (skin 3)
Share-VDE (Bib/Holding dataset)
Sapientia (Share Cluster Knowledge Base)
Share Catalogue
External sources (VIAF, ISNI, LCSH, FAST)
Share-ART (Bib/Holding dataset)
Share-MUSIC
Share-PCC
Share-VDE tenant
Share-ART tenant
Share-MUSIC tenant
Share-PCC tenant
Share Catalogue tenant

Inclusion of the PCC in Share-VDE

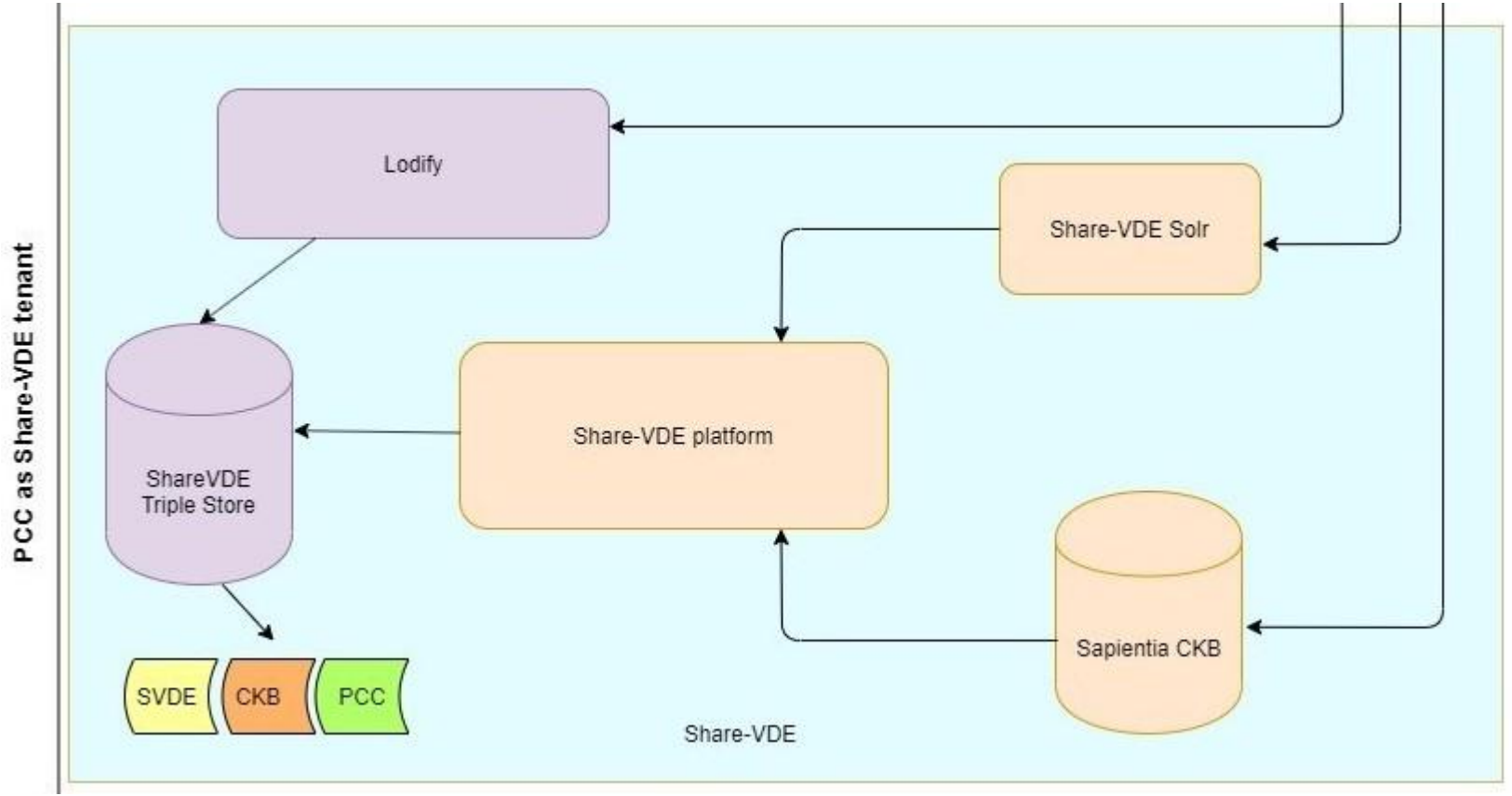
Inclusion of PCC data into the Share initiative as autonomous tenant with a separated enriched CKB and local PCC URIs



Inclusion of PCC data into Share-VDE as tenant: first step



Inclusion of PCC data into Share-VDE as tenant: second step



Workflow PCC - Share-VDE

1. The PCC can be considered as a new library to be included in Share-VDE workflow;

2. the PCC MARC data feed the Share-VDE clusterization tool Authify;

3. enrichment of the PCC MARC data through Authify;
    - URI enrichment from various data sources: LC, GND, VIAF, OCLC, Wikidata, Share-VDE;
    - SVDE is an external data source serving URI enrichment for PCC;
4. result:
    - “ad hoc” PCC Cluster Knowledge Base including SVDE, LC, GND, VIAF, OCLC, Wikidata URIs (and URIs from other sources);
    - the PCC MARC data are data enriched;

Workflow PCC - Share-VDE

5. the PCC MARC data enriched would feed
    - Sapientia (i.e. the SVDE Cluster Knowledge Base);
    - SVDE platform (through SOLR);
    - Lodify conversion tool (conversion from MARC to linked data).

6. PCC data would have a specific URI to identify the entities of PCC graph, e.g.
   http://pcc.data.gov/rdfBibframe/Agent/1234567
   SameAs
   http://share-vde.org/sharevde/rdfBibframe/Agent/38006
   SameAs
   http://isni.org/isni/0000000121032683

7. PCC data would be enriched with Share-VDE URIs (and other sources).

Benefits of having PCC data in Share-VDE

Transition to linked data:

PCC data converted in RDF (both new data and old records)

original MARC records enriched with URIs

Persistent identification and authoritativeness:

enrichment of PCC data with IDs from authoritative sources (SVDE, ISNI, VIAF etc.)

ad hoc PCC namespace for PCC URIs → this makes the PCC itself an authoritative source

enhanced outreach of the PCC and prominent role in the international community

Exchange of information:

easier integration of PCC data in external environments through the PCC URIS (e.g. Wikidata)

What comes next

- Initial load from OCLC to SVDE by 31st December 2020 → approx. 4.5 million WorldCat MARC bibliographical records created by PCC libraries (042 ##$apcc)
- SVDE will deliver PCC records converted in BIBFRAME along with the original MARC records enriched with URIs
- to follow, PCC entities will be available in the CKB
- PCC data on SVDE portal and/or on PCC dedicated skin
- after the initial load, regular updates from OCLC are foreseen, on monthly basis, for the duration of LD4P3

Thank you

https://share-vde.org
info@share-vde.org
https://wiki.share-vde.org/